SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 11, No. 4, (2025), pp. 3532-3545
Open Access. Online & Print

DOI: 10.5281/zenodo.11425178

AI-BASED PROTOTYPE FOR TRAINING THE VENIPUNCTURE SURGICAL PROCEDURE

Oscar E. Gualdrón-Guerrero[1], Johan D. Florez Vera[2], Marisol Maestre Delgado[3]*

[1]A&C Research Group, Faculty of Engineering and Architecture, Electronic Engineering, Universidad de Pamplona, Km 1 vía Bucaramanga, 543050, Pamplona, Colombia
Agroinnova Research Group, Ernesto Schiefelbein Foundation, Bogotá, Colombia
[2]A&C Research Group, Faculty of Engineering and Architecture, Electronic Engineering, Universidad de Pamplona, Km 1 vía Bucaramanga, 543050, Pamplona, Colombia
[3]Faculty of Economic and Business Sciences, Universidad de Pamplona, Pamplona, Colombia, Grupo de Investigación Gramy, Email: marisol.maestre@unipamplona.edu.co

Received: 29/06/2025
Accepted: 03/11/2025

Corresponding Author: Marisol Maestre Delgado (marisol.maestre@unipamplona.edu.co)

ABSTRACT

The article describes the development of a prototype that uses artificial intelligence to train in the venipuncture procedure. This system combines a wearable device with sensors and an anatomical forearm prosthesis, designed to simulate five different clinical conditions. These conditions seek to replicate real-life scenarios that healthcare professionals must learn to accurately manage. The prototype's software includes an artificial intelligence model that classifies the user's performance in real time, labeling them as expert or novice. This approach aims to improve the learning of a key skill in clinical settings, reducing risks for patients and providing an objective assessment of the user's progress. The results obtained suggest that this technology represents an effective alternative to strengthen teaching and clinical practice, all in a safe and controlled environment. This advancement could transform the way healthcare professionals train, ensuring more accurate and safe learning.

KEYWORDS: Artificial Intelligence; Wearable Device; Venipuncture; Surgical Simulation; Medical Training; Clinical Skills Assessment.

Copyright: © 2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License. (https://cre-ativecommons.org/licenses/by/4.0/).

1. INTRODUCTION

Venipuncture is a fundamental procedure in clinical practice (Pham, Aldous, & Brincat, 2024), used for the extraction of blood samples for diagnostic purposes, as well as for the timely administration of medications or therapeutic tests. Although it is a routine technique, incorrect execution can pose risks such as local infections, hematomas, or vascular injuries. Therefore, it requires clinical personnel to have an adequate skill level to perform the insertion safely and accurately (Vegas et al., 2025), ensuring patient safety and minimizing associated complications.

However, training in this technique faces significant challenges due to its invasive nature and the need for technical precision. Among the main obstacles are the limited availability of opportunities for supervised practice, the risks involved in learning on real patients, and the subjective evaluation of performance by instructors (Pedrett et al., 2022; Hackney, O'Neill, O'Donnell, & Spence, 2023).

Moreover, various factors contribute to the complexity of the procedure, such as patients' anatomical variability, the presence of small or fragile veins, obesity, history of previous treatments, and operator inexperience. These elements reduce the first-attempt success rate and, when multiple attempts are required, increase the likelihood of adverse effects such as infections or local trauma, which can seriously compromise the patient's health (Pepley et al., 2017).

In recent years, the integration of advanced technologies such as artificial intelligence (AI) and wearable devices has shown enormous potential for developing healthcare applications. This field, known as mobile health or mHealth, refers to the use of mobile and wireless technologies to improve access to, monitoring of, and quality of healthcare (Alòs & Puig-Ribera, 2021).

According to (Bajwa, Munir, Nori, & Williams, 2021), the increasing availability of multimodal data (including clinical, demographic, economic, and physiological information), along with innovations in the Internet of Things (IoT), digital security, and mobile telephony, is driving the convergence between healthcare and technology. This progress paves the way for the implementation of intelligent systems that can transform traditional service delivery models by integrating AI algorithms to support diagnosis, treatment, and clinical training.

The study presented in (Ávila-Tomás, Mayer-Pujadas, & Quesada-Varela, 2021) identifies multiple applications of AI in the healthcare field, ranging from prevention and early diagnosis to therapeutic monitoring and medical education. Particularly, it highlights the use of AI as a support tool in the continuous training of healthcare professionals, enabling the creation of controlled simulated environments, personalized feedback, and quantitative evaluation of student performance. In this regard, AI helps reduce instructor subjectivity by providing accurate and repeatable metrics for analyzing learners' progress (Anthamatten & Holt, 2024).

The perspective found in the literature suggests that the use of smart technologies in medical education can accelerate the acquisition of clinical skills, optimize teaching-learning processes, and allow users to develop technical competencies more effectively and safely, especially in high-frequency procedures such as venipuncture.

One of the main advantages of simulation in educational contexts (Saleem & Khan, 2023) is that it allows students to practice without putting patients at risk. This approach has gained importance over the past decade, in parallel with the growth of research aimed at improving the quality of invasive technique training (Velásquez-Salazar, Gaitán-Buitrago, & Becerra-Cardona, 2024).

In this context, artificial intelligence has emerged as a key tool in transforming medical education (Silcox et al., 2024), facilitating not only teaching but also the objective assessment of clinical skills (Alowais et al., 2023). Recent studies (Pérez, 2024; Keri et al., 2022) have shown that AI-based systems can provide real-time feedback and adjust the complexity of training according to the student's level, thereby optimizing the learning experience.

Additionally, combining artificial intelligence with wearable devices enables the real-time collection and analysis of biometric data during the execution of the procedure (Karako, 2024; Lu et al., 2020), which supports objective performance evaluation and allows continuous monitoring, even in the absence of direct supervision.

Regarding recent advances in the specific area of venipuncture, intelligent systems have been developed that integrate computer vision, sensors, and machine learning algorithms. For example, the VeniBot system (Chen et al., 2021) proposes a compact robotic approach that uses deep learning for semi-supervised segmentation of ultrasound images, facilitating the localization of hard-to-access veins.

Another notable development is presented in (Salcedo & Peñaloza, 2023), which introduces a portable system for venipuncture assistance in the antecubital fossa. It uses near-infrared imaging and a

Raspberry Pi 4B as the base platform. This device employs neural networks to enhance vein visibility in patients with complex vascular access.

In (Liu, Petersen, & Zhang, 2020), a method is introduced to automatically segment the phases of the needle insertion procedure using sensors that record motion and force data. The system differentiates between phases such as approach, skin penetration, and final insertion, and has been tested with students to distinguish between different skill levels, showing promising results.

This work proposes the development of a prototype that combines wearable technology and artificial intelligence for objective training in venipuncture. The system allows students and professionals to practice the procedure in a controlled environment, recording the user's biomechanical signals in real time during needle insertion.

The prototype includes a forearm prosthesis that simulates various clinical conditions (Welyczko, 2020), such as healthy veins, fragile veins, edema, narrow or fistulized vascular paths, offering realistic and challenging scenarios that reflect common situations in clinical practice. This provides a risk-free environment for both the patient and the practitioner.

On the software side, the system incorporates an artificial intelligence model trained to classify user performance based on their experience level, distinguishing between experts (trained doctors and nurses) and trainees (students), in order to provide objective feedback and traceability of progress.

The results obtained suggest that the system represents a viable and effective tool for strengthening venipuncture training and assessment processes, contributing to educational quality improvement and the safe preparation of future healthcare professionals.

2. MATERIALS AND METHODS

The study conducted is classified as quasi-experimental (Bärnighausen et al., 2017), with a comparative approach.

The main objective was to evaluate the effectiveness of a system based on sensors and artificial intelligence (AI) algorithms to classify performance in the venipuncture procedure between two groups of participants: experts and non-experts.

To achieve this, data were collected using a wearable wristband equipped with accelerometers, gyroscopes, and pressure sensors, which were used to capture time series (Hu, Sun, Li, & Zhang, 2023) during the performance of the venous cannulation procedure. The time series data from the procedures were processed to train an artificial intelligence algorithm with the primary goal of classifying participants' performance according to their level of experience. The analysis focused on comparing the classifier's results in distinguishing between the two groups.

2.1 Prototype design

The prototype design consisted of the integration of hardware and software to simulate a controlled clinical environment, allowing healthcare professionals to practice venipuncture procedures without posing any risk to real patients. The device was designed to capture real-time data during the procedure and to classify the user's skill level through the training of an artificial intelligence model.

2.1.1 Forearm prosthesis for venipuncture

For this study, prosthetic models provided by the Canadian company ProxSIMity Inc. were used. These simulate five distinct clinical conditions for venipuncture training: healthy, edema, fragile, thin veins, and fistula-graft. These variations allow healthcare trainees to practice in scenarios with different levels of difficulty, enhancing their vascular access skills.



Figure 1. Forearm prosthesis support. Source. ProxSIMity.



Figure 2. Forearm prosthesis, healthy condition. Source. ProxSIMity.



Figure 3. Forearm prosthesis, fragile condition. Source. ProxSIMity.

Figure 1 shows the support structure for the forearm prostheses, while Figures 2 and 3 present samples of the prostheses for the healthy and fragile conditions, respectively.

The conditions are described as follows:

- Healthy: Represents well-conditioned tissue, with visible and palpable veins that facilitate venipuncture.
- Edema: Simulates swelling on the back of the hand, making vein palpation and visualization more difficult.
- Fragile: A condition where the skin and tissues are delicate, increasing the risk of vascular damage.
- Thin veins: Represents narrow and less visible veins, making localization and access more challenging.
- Fistula-graft: Simulates an artificial connection between an artery and a vein, typically found in patients with renal failure, requiring special care to avoid injury.

The venipuncture kit incorporates a quick-swap design to easily switch between different conditions. The replaceable units are designed to withstand hundreds of punctures. Once the veins become worn, only the individual units need to be replaced, making the simulator both cost-effective and easy to maintain.

2.1.2 Description of the wearable device

The wearable device used in this research is a wristband equipped with an accelerometer, a gyroscope, and pressure sensors located on the thumb and index finger, as shown in Figure 4. This device, provided for the study by the Canadian company ProxSIMity Inc., enables real-time data collection during the surgical procedure. Through its sensors, it records variations in tilt across the X, Y, and Z axes, acceleration of movements along these axes, and the pressure applied by the index finger and thumb throughout the process.



Figure 4. Wearable device with integrated sensors. Source. ProxSIMity.

The wristband connects via Bluetooth technology (Verma, Singh, & Kaur, 2015) to transmit the data to a computer, where it is stored in .csv format for subsequent analysis.

To manage this connection, a custom software application was developed using Unity (Unity Technologies, 2025). The interface allows for managing the Bluetooth connection with the wristband, initializing data collection, pausing acquisition, and saving the recorded data, as shown in Figure.



Figure 5. Connection interface functionalities. Source. Own work.

2.2 Experimental procedure

2.2.1 Data acquisition and labeling

Specific venipuncture procedure tests were carried out with the objective of collecting comparative data between two groups of participants—experts and trainees—in order to later train the artificial intelligence model. The participants were divided as follows:

- Expert group: Physicians, nurses, and faculty from the medicine, nursing, and related programs at the University of Pamplona.
- Trainee group: Students from the medicine, nursing, and related programs at the same institution.

A total of 16 individuals participated, evenly distributed between the two groups (8 experts and 8 trainees). Each participant performed 5 venous cannulation trials, resulting in a total of 80 samples: 40 from experts and 40 from trainees. This ensured class balance for the subsequent training stage of the AI model. All samples were stored and labeled

accordingly into their respective categories during the data acquisition process.

For data collection, the sensor-equipped wristband was worn by the participant while performing the procedure. Additionally, the prosthesis was prepared, and detailed instructions on the standardized protocol were provided, consisting of the following steps:

a. Needle delivery: The participant was handed the appropriate needle for performing venipuncture on the prosthesis.

b. Vein location: Participants were instructed to carefully palpate the prosthesis and locate the simulated venous path before proceeding.

c. Start of data recording: Once ready, the participant pressed a button on the sensor wristband to start data capture.

d. Venipuncture execution: The participant performed the venipuncture at the previously identified site, simulating the clinical venous access procedure.

e. Stop of data recording: Upon completion of the needle insertion, the participant pressed the button again to stop signal capture.

f. Success verification: Success was verified by extracting fluid through the simulated venous system.

g. Procedure labeling: Finally, the procedure result was labeled via the data acquisition software interface, registering whether the attempt was successful or not.

2.2.2 Exploratory data analysis.

The data collected during the venous cannulation tests were organized and processed using a Pandas DataFrame. This dataset included:

- Accelerations: The three components (X, Y, Z) recorded by the accelerometer, capturing the participant's linear movements during the procedure.
- Gyroscope angles: The roll, pitch, and yaw values describing hand orientation and rotation during execution.
- Pressure signals: Measurements from the pressure sensors on the thumb and index finger, registering the applied force while handling the catheter.

After loading the data, an integrity check was performed to ensure no missing or inconsistent values. Time series plots were then generated for each variable to visually explore their behavior.

Figures 6, 7, and 8 show sample signals captured by the prototype's sensors. The graphs display tilt on the X-axis, acceleration on the X-axis, and pressure readings from the thumb sensor, respectively.

Each graph shows 16 signals, representing one of the five repetitions performed by each participant during testing. Signals shown in blue correspond to individuals classified as "Experts," while those in red represent individuals classified as "Novices".

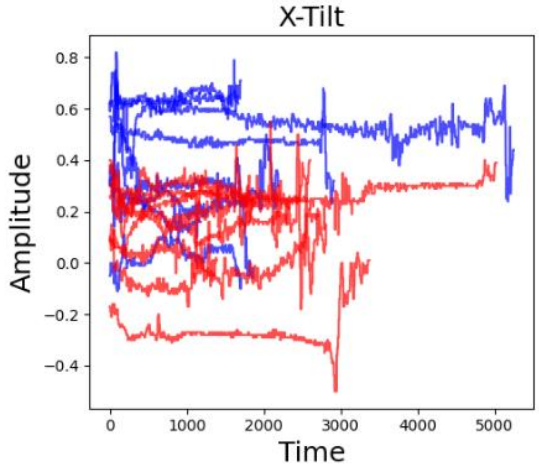


Figure 6. Sample signals from the tilt sensor on the X-axis. Source. Own work.

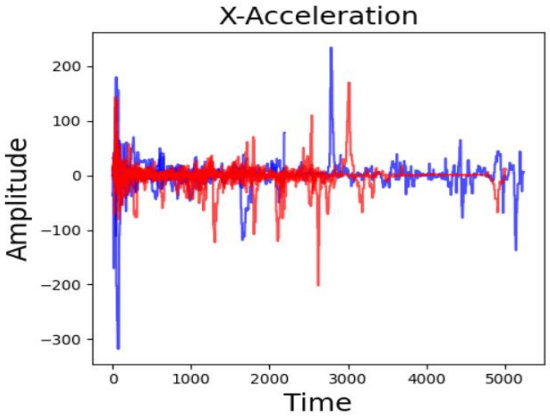


Figure 7. Sample signals from the acceleration sensor on the X-axis. Source. Own work.

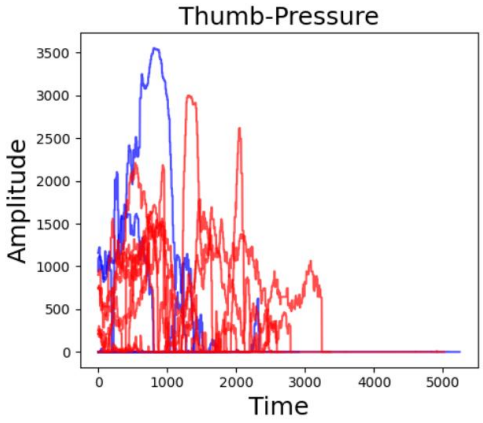


Figure 8. Sample signals from the thumb pressure sensor. Source. Own work.

The time series graphs reveal a high degree of variability in the data, even among samples within

the same category, making it extremely difficult to visually identify clear patterns or differences between experts and novices.

These characteristics in the time series highlight the need for more advanced processing and analytical techniques—such as feature extraction and the use of artificial intelligence models—to extract meaningful information and build an effective classification model. It also became evident that the collected samples varied in length, due to differences in the duration of each procedure depending on the participant's skill, experience, and execution pace. This variability poses a significant challenge for developing AI models, as most machine learning algorithms require input data to have a uniform and consistent structure.

This requirement stems from the fact that AI algorithms—especially those based on neural networks or supervised learning methods—typically assume a fixed input dimensionality. When samples have different lengths, additional preprocessing is required to adapt them into a compatible format, which may introduce distortions or result in the loss of important information. Regarding variable-length time series, reference (Bier, Jastrzębska, & Olszewski, 2022) notes that textbooks on multivariate time series classification problems rarely address the challenge posed by differences in time series lengths. However, this issue is common in real-world applications. The study proposes three approaches to handle this situation: truncation, data padding, and missing value estimation.

This problem is also thoroughly addressed in (Kazijevs & Samad, 2023), with a particular focus on multivariate time series. The study explores various data imputation methods, including deep learning techniques, applied to five health-related datasets. The authors conduct a detailed review of recent methods for imputing multivariate time series, focusing on deep learning-based approaches published within the last five years. Finally, this variability in signal duration may mask key patterns that differentiate experts from novices, since differences in execution speed and procedure length may not directly reflect the participant's skill level, but rather other factors such as caution, personal technique, or even task complexity. As such, uneven sample lengths complicate the technical processing of data and may impact the model's ability to generalize and learn effectively.

2.2.3 Feature extraction

After obtaining the dataframes, statistical and signal-based features (Krishnan & Athavale, 2018) were extracted for each sensor using Python's NumPy library (Harris et al., 2020).

This method is widely used in time series analysis (Wang et al., 2022), especially in biomedical applications. In such cases, signals are often noisy and complex; therefore, feature extraction helps reduce noise and highlight relevant patterns (Rajoub, 2020). This process transforms raw signals into a set of numerical descriptors that summarize their temporal, statistical, or frequency-related behavior, making it easier to identify meaningful patterns.

The issue discussed in the previous section regarding signal length variability was addressed through feature extraction, which allowed these differences to be normalized.

By extracting statistical and time-domain features, a more homogeneous data representation is obtained. This reduces the influence of individual variations and enables classification algorithms to focus on the relevant signal patterns. As a result, improved model accuracy is expected.

In this case, the features extracted are listed in Table 1:

Tabla 1. Extracted features.

Name	Equation	Meaning or Application
Mean	$\mu = (1/n) \sum x_i$	Average of the values; represents the central tendency.
Standard deviation	$\sigma = \sqrt{[(1/n) \sum (x_i - \mu)^2]}$	Measures the spread of the data relative to the mean.
Mean absolute deviation (MAD)	$MAD = (1/n) \sum \lvert x_i - \mu \rvert$	Average absolute deviations from the mean.
Minimum value	min(x)	Lowest value in the signal.
Máximum value	max(x)	Highest value in the signal.
Range (max-min)	max(x) - min(x)	Total range of the signal.
Median	mediana(x)	Central value of the ordered signal; less sensitive to outliers.
Median absolute deviation (MAD_med)	$MAD_med = (1/n) \sum \lvert x_i - \text{mediana}(x) \rvert$	Dispersion with respect to the median.
Interquartile range (IQR)	IQR = Q3 - Q1	Dispersion between the first and third quartile.
Count of negative values	$\sum (x_i < 0)$	Number of values below zero.
Count of positive values	$\sum (x_i > 0)$	Number of values above zero.
Count of values above the mean	$\sum (x_i > \mu)$	Frequency of values greater than the mean.
Number of peaks	peaks(x)	Number of local maxima; useful for analyzing repetitive patterns.
Skewness	skew(x)	Degree of symmetry in the signal distribution.
Kurtosis	kurt(x)	Concentration level of values relative to the mean.
Energy	$E = \sum x_i^2$	Measure of the signal's total power.

Source. Own work.

Each of these features was calculated individually for every sensor: tilt in the X, Y, and Z axes; acceleration in the X, Y, and Z axes; and pressure on the thumb and index finger.

This process yielded eight independent dataframes, one for each sensor, as shown in Figure 9. Each dataframe has dimensions of 80×16 (80 trials and 16 features). Additionally, a consolidated dataframe was generated, combining all sensors, with dimensions of 80×128.

2.2.4 Application of principal component analysis (PCA).

To reduce data dimensionality while preserving as much information as possible, Principal Component Analysis (PCA) was applied to each of the dataframes obtained after feature extraction. The use of PCA facilitates data visualization, accelerates the training process of classification models, and helps mitigate problems related to multicollinearity or redundancy among variables.

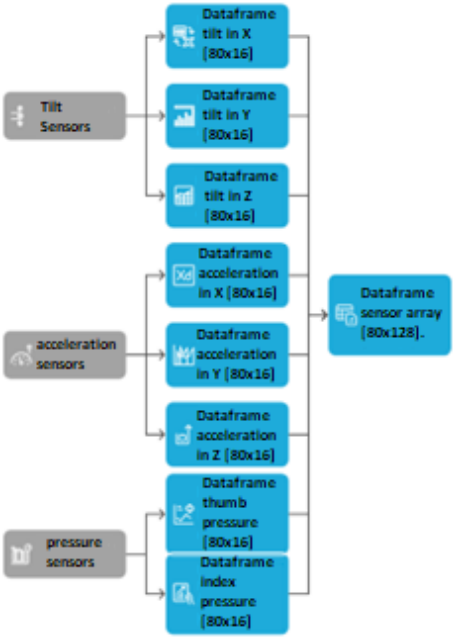


Figure 9. Resulting dataframes from the feature extraction process. Source. Own work.

The effectiveness of PCA for dimensionality reduction and improved classification performance has been widely documented in the scientific literature. For example, in study (Tăuțan, Rossi, de Francisco, & Ionescu, 2020), the application of PCA to EEG signals for automatic sleep stage detection was shown to maintain or even enhance classification model accuracy while reducing computational load. Similarly, in (Rajagopal & Ranganathan, 2016), the performance of linear dimensionality reduction techniques, including PCA, was evaluated for cardiac arrhythmia classification, showing that PCA helped improve both sensitivity and F-score metrics of the models.

Through this technique, the original set of 16 features per sensor was projected into a new lower-dimensional space composed of two principal components, selected based on their ability to preserve the highest possible proportion of total data variance. This procedure was implemented using the PCA function from Python's sklearn.decomposition module (Géron, 2019). As a result, nine dataframes were obtained—each corresponding to a different sensor—with dimensions of 80×2 (80 samples and 2 principal components).

Table 2 presents the percentage of variance explained by the first two principal components obtained through PCA for each sensor. It reports the individual values for the first and second components, as well as the cumulative percentage of total variance explained by both. This information is essential for evaluating the effectiveness of the dimensionality reduction process, ensuring that the selected components retain most of the relevant information for the subsequent classification task.

Table 2. ercentage of variance explained by the first two principal components, by sensor.

Sensor	Component 1 (%)	Component 2 (%)	Total variance (%)
Tilt in X	74,13	22,83	96,96
Tilt in Y	81,65	14,53	96,18
Tilt in Z	74,5	20,55	95,05
Acceleration in X	57,46	28,12	85,58
Acceleration in Y	79,55	13,05	92,6
Acceleration in Z	73,87	15,79	89,66
Thumb pressure	100,00	0,00	100,00
Index pressure	100,00	0,00	100,00
Sensor array	80,32	19,68	100,00

Source. Own work.

As shown in Table 2, in all cases the first two components retain between 85% and 100% of the original variance, justifying their use as a reduced representation of the data without significant information loss.

In conclusion, as shown in Figure 10, a total of 18 dataframes were generated: nine with the original extracted features from each sensor and nine with their respective reduced representations obtained through PCA. These datasets were used as input for the classification models evaluated in this study.

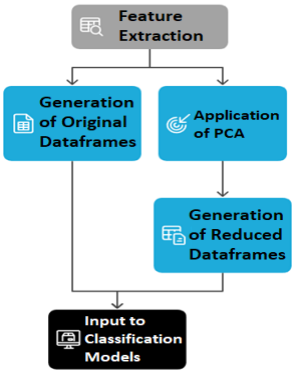


Figure 10. Consolidated DataFrames. Source. Own work.

2.2.5 Artificial intelligence models.

The objective of the artificial intelligence model was to develop a system capable of classifying users according to their level of experience in performing the clinical venipuncture procedure, distinguishing between expert users and trainees (novices). To achieve this, two supervised learning models were implemented. Based on the data collected through wearable wristband-integrated sensors—and following preprocessing, structuring into dataframes, and sample labeling—a Multi-Layer Perceptron (MLP) neural network and a Support Vector Machine (SVM) were trained and evaluated. The Multi-Layer Perceptron (MLP) is an artificial neural network architecture composed of multiple hidden layers arranged sequentially between the input and output layers. Each neuron in these layers applies nonlinear activation functions, enabling the MLP to learn and represent complex, nonlinear relationships between input features (Li et al., 2022; Weiss, Karimijafarbigloo, Roggenbuck, & Rödiger, 2022; Bikku, 2020).

This structure facilitates the identification of hidden patterns and the approximation of functions beyond the capabilities of traditional linear techniques, making MLP a valuable tool in fields such as speech recognition, computer vision, bioinformatics, and biomedical data analysis. Figure 11 shows a general schematic architecture of a Multi-Layer Perceptron (MLP) network, illustrating the approach used to classify users as experts or novices during the venous cannulation procedure. It depicts the typical arrangement of the network layers, as well as the types of input data used: biomechanical signals from tilt, acceleration, and finger pressure sensors, restructured as described in the previous section.

On the other hand, Support Vector Machines (SVMs) are widely used supervised learning algorithms in biomedical and health-related Big Data analysis for classification and regression tasks. Their ability to handle high-dimensional data and detect complex patterns makes them a powerful tool in bioinformatics, clinical medicine, and health data analysis (Taha, 2025). SVM models (Azizi, 2024; Cervantes, Garcia-Lamont,

Rodríguez-Mazahua, & Lopez, 2020) work by finding a hyperplane that maximally separates the different classes in a feature space. For nonlinear data, kernel functions are used to project the data into a higher-dimensional space where linear separation becomes possible. This makes the algorithm effective for solving complex problems in areas such as bioinformatics and pattern recognition.

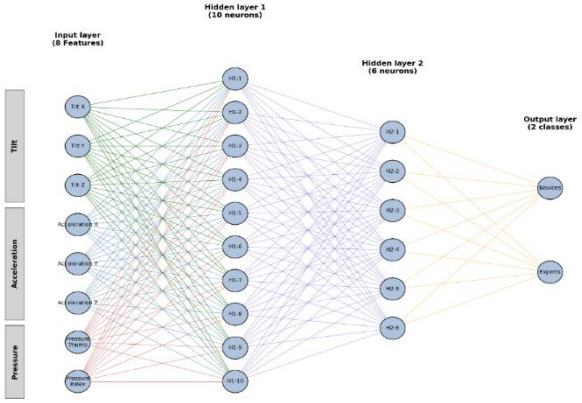


Figure 11. General MLP architecture for classifying novice and expert users. Source. Own work.

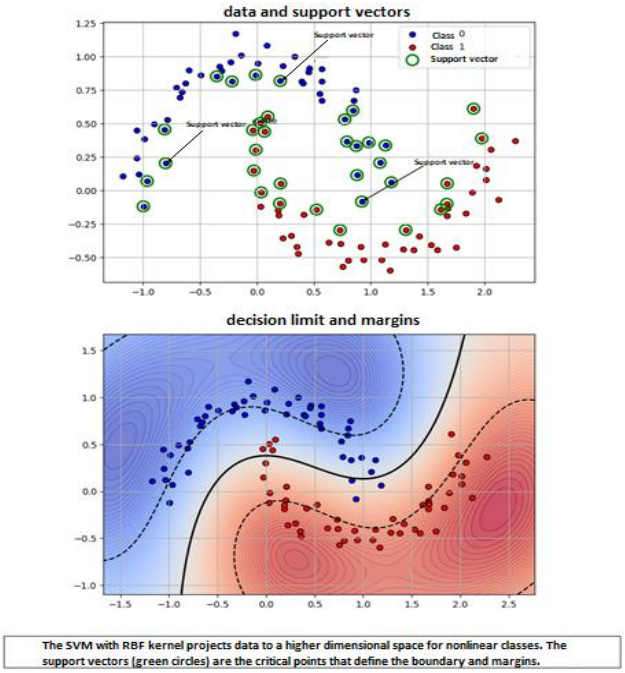


Figure 12. Illustration of a Support Vector Machine (SVM) with RBF kernel for nonlinear classification. Source. Own work.

In the first panel, samples from two different classes are shown, with green circles highlighting the support vectors—the critical data points that define the decision

boundary. The second panel displays the optimal decision boundary (solid black line) obtained by the model, as well as the margins (dashed black lines) that maximize the separation between the classes.

The shaded regions represent the areas of the feature space assigned to each class by the classifier. This visualization illustrates the SVM's ability to transform the original space through kernel functions, enabling nonlinear class separation and emphasizing the essential role of support vectors in defining the optimal hyperplane.

2.2.5.1 MLP neural network specification.

The architecture of the designed MLP network is shown in Figure 13. It consists of an input layer that accepts 2, 16, or 128 features, depending on the dataset used: 2 features in the dataframes processed using Principal Component Analysis (PCA), 16 features in those derived from individual sensor feature extraction, and 128 features when using the full consolidated sensor dataframe.

Two hidden layers follow the input layer. The first hidden layer contains 50 neurons, and the second has 25 neurons. Both layers use the ReLU (Rectified Linear Unit) activation function (Bai, 2022), which introduces nonlinearity into the model and facilitates the learning of complex relationships between input variables. The output layer consists of two neurons, corresponding to the "expert" and "novice" classes, and employs the softmax activation function (Almurieb & Bhaya, 2020), which converts the outputs into a probability distribution over the two classes.

The model was compiled using the binary cross-entropy loss function (Mao, Mohri, & Zhong, 2023) and optimized using the Adam algorithm (Hospodarskyy et al., 2024), which adjusts the network's weights to minimize the loss function. The selected evaluation metric was accuracy. Training was carried out over 150 epochs with a batch size of 10, meaning that the model's weights were updated after processing each subset of 10 samples.

This MLP configuration was specifically designed to effectively tackle the classification task, enabling the model to identify complex patterns and accurately distinguish between different levels of experience in the execution of the venipuncture procedure.

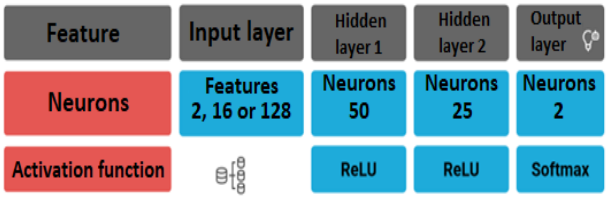


Figure 13. Architecture of the designed MLP neural network. Source. Own work.

2.2.5.2 SVM model specification.

The Support Vector Machine (SVM) model implemented in this study uses the Radial Basis Function (RBF) kernel (Thurnhofer-Hemsi, López-Rubio, Molina-Cabello, & Najarian, 2020), a widely used technique for addressing nonlinear classification problems. This approach calculates a Gaussian function that evaluates the similarity between pairs of data points by measuring their distance relative to a reference center. The RBF kernel allows the input data to be transformed into a higher-dimensional feature space, where previously non-separable classes can become linearly separable.

2.2.6 Classifier model performance evaluation.

To evaluate the performance of the classification models applied to the venous cannulation procedures, accuracy was used as the primary evaluation metric. The models were validated using both the entire training dataset and a specific data split to ensure a more robust evaluation. A summary of this process is presented in Figure 14.

For the MLP model, the dataset was split into 60% for training and 40% for validation. For the SVM model, a 70% training and 30% validation split was used. This partitioning methodology was systematically applied to each of the 18 dataframes previously described, with the goal of identifying which sensors provide the highest informational value for the evaluated processes. The analysis also included the exclusion of sensors that, based on the results, contributed less information. In particular, the study evaluated whether removing these signals from the "sensor feature set" dataframe would positively or negatively affect the accuracy percentage of the classification models. This strategy aimed to optimize system performance by reducing dimensionality without compromising the model's precision.

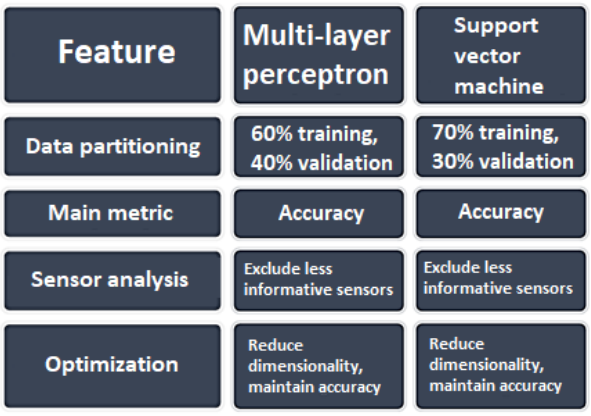


Figure 14. Summary of the model performance evaluation process. Source. Own work.

3. RESULTS

3.1 Classifier model performance by sensor.

To evaluate the performance of the classification models in the emulated venous cannulation procedure, accuracy rates were analyzed after training the MLP and SVM models on training and testing sets using the data from each individual sensor, as well as from sensor combinations. Both the features obtained through the extraction process and those reduced via Principal Component Analysis (PCA) were considered. The results are summarized in Figures 15, 16, and 17.

In general, acceleration and pressure sensors showed greater discriminative power than tilt sensors when evaluated individually. The acceleration sensor on the X-axis achieved outstanding performance with the SVM model, reaching 100% accuracy in the training set and 58% in the test set, while with the MLP model, it achieved 95% accuracy in simple validation. These results highlight its ability to capture significant variations during the procedure. In the case of pressure sensors, both the thumb and index finger sensors yielded competitive results. With SVM, both achieved 100% accuracy in training and 62% in testing. With MLP, the highest values were observed in simple validation: 76.25% for the thumb sensor and 67.5% for the index finger sensor. This suggests that the applied forces during puncture are useful indicators for classifying user experience level.

Conversely, tilt sensors, particularly on the Z-axis, showed more limited classification capability. With MLP, accuracy was 61.25% in simple validation and only 53.13% in partition-based validation; with SVM, although 100% training accuracy was observed, performance dropped to 54% or less in testing, indicating overfitting and poor generalization.

However, after applying dimensionality reduction using PCA, an improvement in performance was observed for some of these sensors. For example, the Z-axis tilt sensor improved from 61.25% to 76.25% accuracy in MLP validation, and from 54% to 62% in SVM testing, highlighting the usefulness of PCA in reducing redundancy and increasing model robustness.

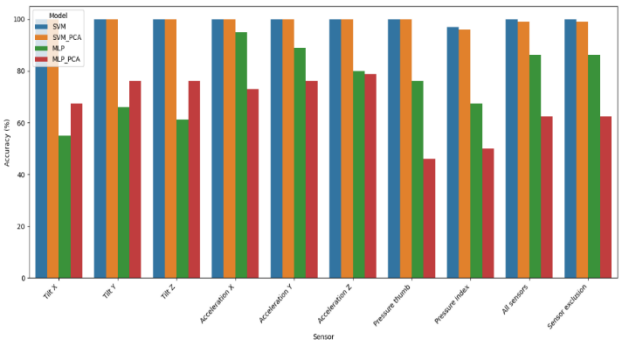


Figure 15. Classifier performance on training set. Source. Own work.

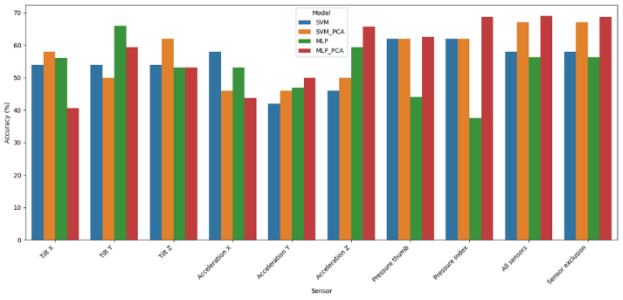


Figure 16. Classifier performance on test set. Source. Own work.

3.2 Sensor combination and selective exclusion.

When all sensors were used in combination, the highest overall performance was achieved in classifying the users' level of experience. With the SVM model, 100% accuracy was obtained on the training set and 58% on the test set using the original features. After applying dimensionality reduction via PCA, the model achieved a very similar training accuracy (99%) and an improved test accuracy of 67%, indicating better generalization through the elimination of signal redundancies.

In the case of the MLP model, performance was also high. Using the original features from all sensors, the model achieved 86.25% accuracy in simple validation and 56.25% in partitioned validation. After applying PCA, simple validation decreased to 62.50%, but partitioned validation increased to 69%, reflecting a positive effect of preprocessing on the model's robustness.

Additionally, a configuration was tested in which less informative sensors (particularly the tilt sensors) were excluded. The results were nearly identical to those obtained with the full sensor set. With MLP, accuracy remained at 86.25% (simple validation) and 56.25% (partitioned validation) without PCA. After PCA, the model again achieved 62.50% and 68.75%, respectively. The SVM model showed virtually the same performance as with the complete configuration, reaching 100% training accuracy and 67% testing accuracy with PCA.

This similarity suggests that excluding low-contribution sensors does not significantly affect the predictive capability of the system and may even contribute to a simplified design without compromising performance.

3.3 Model comparison

When comparing the behavior of the two models used, it was observed that the SVM model showed greater stability between training and testing phases, especially in multi-sensor configurations. In most combinations, the model achieved 100% accuracy on the training set, while on the test set it reached values between 46% and 67%, depending on the sensor or combination used. Notably, when combining all sensors with PCA, SVM achieved its highest test accuracy: 67%.

On the other hand, the MLP model showed greater variability in its results, with marked differences between simple and partitioned validations. Nevertheless, competitive results were achieved in several configurations. For example, using all sensors without applying PCA, the model reached 86.25% accuracy in simple validation. When excluding the less informative sensors and applying PCA, MLP achieved its highest partitioned validation accuracy: 68.75%.

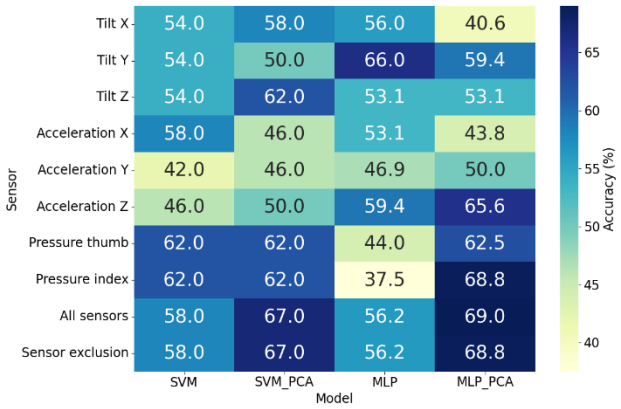


Figure 17. Heatmap of accuracy in the test set. Source. Own work.

4. DISCUSSION

The results obtained in the venipuncture emulation procedure highlight the potential of artificial intelligence models—particularly Support Vector Machines (SVM)—to accurately distinguish between different levels of expertise in medical simulation environments. The high accuracy achieved by this model, especially when using multiple sensors and selectively excluding low-performing ones, underscores the importance of appropriate sensor selection and data fusion in building reliable automated evaluation systems.

From a medical education perspective, these findings are highly relevant. Venipuncture is a critical clinical skill, particularly in emergency contexts, where technical competence directly impacts patient safety. The implementation of intelligent systems capable of providing objective, continuous, and bias-free assessments of trainees represents a significant advancement over traditional evaluation methods that rely on subjective observation by instructors.

Compared to existing training and assessment methods—often based on qualitative rubrics or checklists—the proposed approach offers automated classification. Furthermore, the integration of wearable sensors and classification models allows for personalized feedback based on individual performance, promoting more efficient and adaptive learning. While previous studies have employed tools such as high-fidelity simulators or video analysis, these solutions tend to be expensive, require constant supervision, and do not always capture real-time kinematic or pressure-related variables.

Nonetheless, the present study has certain limitations. First, while the sample used is representative, expanding the dataset would improve model robustness and generalizability. Second, the use of dimensionality reduction techniques such as PCA showed mixed effects, particularly in the MLP model, suggesting the need to explore more advanced feature selection methods or non-linear dimensionality reduction techniques. Additionally, the simulated environment may not fully replicate real-world clinical conditions. Future research should consider evaluations in hospital scenarios using standardized patients or more anatomically realistic models.

Going forward, it will be essential to explore more sophisticated AI approaches beyond classical models such as SVM and MLP. Algorithms based on deep neural networks, such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), or Transformers, may be better suited to capture the temporal dynamics of hand movements and pressure patterns during the procedure.

5. CONCLUSIONS

This study demonstrated the feasibility of using artificial intelligence techniques combined with wearable sensors to objectively and automatically classify clinical skills during venipuncture procedures in simulated settings.

In the context of medical education, these results represent progress toward more objective, scalable, and personalized assessment systems that can

complement traditional clinical training methods. The integration of sensors and intelligent models not only allows for the assessment of expertise but also promotes autonomous and adaptive learning. Finally, it is recommended to continue this line of research by testing deep learning algorithms, expanding the participant pool, and evaluating the system in real clinical environments.

REFERENCES

Almorieb, H. A., & Bhaya, E. S. (2020). SoftMax neural best approximation. IOP Conference Series: Materials Science and Engineering, 871(1), 012040. https://doi.org/10.1088/1757-899X/871/1/012040

Alòs, F., & Puig-Ribera, A. (2021). Uso de wearables y aplicaciones móviles (mHealth) para cambiar los estilos de vida desde la práctica clínica en atención primaria: una revisión narrativa. Atención Primaria Práctica, 3(S1), 100122. https://doi.org/10.1016/j.appr.2021.100122

Alowais, S. A., et al. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. BMC Medical Education, 23(1), 689. https://doi.org/10.1186/s12909-023-04698-z

Anthamatten, A., & Holt, J. E. (2024). Integrating artificial intelligence into virtual simulations to develop entrustable professional activities. The Journal for Nurse Practitioners, 20(9), 105192. https://doi.org/10.1016/j.nurpra.2024.105192

Ávila-Tomás, J. F., Mayer-Pujadas, M. A., & Quesada-Varela, V. J. (2021). La inteligencia artificial y sus aplicaciones en medicina II: importancia actual y aplicaciones prácticas. Atención Primaria, 53(1), 81–88. https://doi.org/10.1016/j.aprim.2020.04.014

Azizi, T. (2024). Comparative analysis of statistical, time–frequency, and SVM techniques for change detection in nonlinear biomedical signals. Signals, 5(4), 736–755. https://doi.org/10.3390/signals5040041

Bai, Y. (2022). ReLU function and derived function review. SHS Web of Conferences, 144, 02006. https://doi.org/10.1051/shsconf/202214402006

Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: Transforming the practice of medicine. Future Healthcare Journal, 8(2), e188–e194. https://doi.org/10.7861/fhj.2021-0095

Bärnighausen, T., et al. (2017). Quasi-experimental study designs series—Paper 4: Uses and value. Journal of Clinical Epidemiology, 89, 21–29. https://doi.org/10.1016/j.jclinepi.2017.03.012

Bier, A., Jastrzębska, A., & Olszewski, P. (2022). Variable-length multivariate time series classification using ROCKET: A case study of incident detection. IEEE Access, 10, 95701–95715. https://doi.org/10.1109/ACCESS.2022.3203523

Bikku, T. (2020). Multi-layered deep learning perceptron approach for health risk prediction. Journal of Big Data, 7, 50. https://doi.org/10.1186/s40537-020-00316-7

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Chen, Y., et al. (2021). VeniBot: Towards autonomous venipuncture with semi-supervised vein segmentation from ultrasound images.

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems (2nd ed.). O'Reilly Media.

Hackney, L., O'Neill, S., O'Donnell, M., & Spence, R. (2023). A scoping review of assessment methods of competence of general surgical trainees. The Surgeon, 21(1), 60–69. https://doi.org/10.1016/j.surge.2022.01.009

Harris, C. R., et al. (2020). Array programming with NumPy. Nature, 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hospodarskyy, O., Martsenyuk, V., Kukharska, N., Hospodarskyy, A., & Sverstiuk, S. (2024). Understanding the Adam optimization algorithm in machine learning. In Proceedings of the 2nd International Workshop on Computer Information Technologies in Industry 4.0 (CITI'2024) (Vol. 3742, Paper 17). CEUR Workshop Proceedings.

Hu, C., Sun, Z., Li, C., & Zhang, Y. (2023). Survey of time series data generation in IoT. Sensors, 23, 6976. https://doi.org/10.3390/s23156976

Karako, K. (2024). Integración de dispositivos portátiles y aprendizaje profundo: nuevas posibilidades para la gestión de la salud y la prevención de enfermedades. Tendencias en Biociencias. https://doi.org/10.5582/bst.2024.01170

Kazijevs, M., & Samad, M. D. (2023). Deep imputation of missing values in time series health data: A review with benchmarking. Journal of Biomedical Informatics, 144, 104440. https://doi.org/10.1016/j.jbi.2023.104440

Keri, S., et al. (2022). Artificial intelligence in medical education: A scoping review. Medical Teacher, 44(2), 134–145. https://doi.org/10.1080/0142159X.2022.2030932

Krishnan, S., & Athavale, Y. (2018). Trends in biomedical signal feature extraction. Biomedical Signal Processing and Control, 43, 41–63. https://doi.org/10.1016/j.bspc.2018.02.008

Li, X.-D., Wang, J.-S., Hao, W.-K., Wang, M., & Zhang, M. (2022). Multi-layer perceptron classification method of medical data based on biogeography-based optimization algorithm with probability distributions. Applied Soft Computing, 121, 108766. https://doi.org/10.1016/j.asoc.2022.108766

Liu, Z., Petersen, L., & Zhang, Z. (2020). A method for segmenting the process of needle insertion during simulated cannulation using sensor data. IEEE Engineering in Medicine and Biology Society Conference Proceedings, 2020, 6090–6094. https://doi.org/10.1109/EMBC44109.2020.9176158

Lu, L., Zhang, J., Xie, Y., Gao, F., Xu, S., Wu, X., & Ye, Z. (2020). Wearable health devices in health care: Narrative systematic review. JMIR Mhealth Uhealth, 8(11), e18907. https://doi.org/10.2196/18907

Mao, A., Mohri, M., & Zhong, Y. (2023). Cross-entropy loss functions: Theoretical analysis and applications. In Proceedings of the 40th International Conference on Machine Learning (Vol. 202, pp. 23803–23828). PMLR.

Pedrett, R., Mascagni, P., Beldi, G., Padoy, N., & Lavanchy, J. L. (2022). Technical skill assessment in minimally invasive surgery using artificial intelligence: A systematic review. medRxiv. https://doi.org/10.1101/2022.11.08.22282058

Pepley, D. F., Gordon, A. B., Yovanoff, M. A., Mirkin, K. A., Miller, S. R., Han, D. C., & Moore, J. Z. (2017). Training surgical residents with a haptic robotic central venous catheterization simulator. Journal of Surgical Education, 74(6), 1066–1073. https://doi.org/10.1016/j.jsurg.2017.06.003

Pérez, J. (2024). La inteligencia artificial como herramienta en la educación médica. Investigación en Educación Médica, 24(5), 45–50. https://scielo.isciii.es/scielo.php?pid=S2014-98322024000500001

Pham, M., Aldous, R., & Brincat, S. (2024). Implementation of ultrasound-guided cannulation training for foundation doctors. Clinical Medicine, 24(6), 100256. https://doi.org/10.1016/j.clinme.2024.100256

Rajagopal, R., & Ranganathan, V. (2016). Critical evaluation of linear dimensionality reduction techniques for cardiac arrhythmia classification. Circuits and Systems, 7(9), 225–234. https://doi.org/10.4236/cs.2016.79225

Rajoub, B. (2020). Characterization of biomedical signals: Feature engineering and extraction. In Biomedical Signal Processing (pp. 29–50). Elsevier. https://doi.org/10.1016/b978-0-12-818946-7.00002-0

Salcedo, E., & Peñaloza, P. (2023). Edge AI-based vein detector for efficient venipuncture in the antecubital fossa. In Advances in Soft Computing (pp. 297–314). Springer. https://doi.org/10.1007/978-3-031-47640-2_24

Saleem, M., & Khan, Z. (2023). Healthcare simulation: An effective way of learning in health care. Pakistan Journal of Medical Sciences, 39(4), 1185–1190. https://doi.org/10.12669/pjms.39.4.7145

Silcox, C., Zimlichmann, E., Huber, K., Rowen, N., Saunders, R., McClellan, M., Kahn, C. N., Salzberg, C. A., & Bates, D. W. (2024). The potential for artificial intelligence to transform healthcare: Perspectives from international health leaders. npj Digital Medicine, 7(1), 88.

Taha, K. (2025). Machine learning in biomedical and health big data: A comprehensive survey with empirical and experimental insights. Journal of Big Data, 12, 61. https://doi.org/10.1186/s40537-025-01108-7

Tăuțan, A.-M., Rossi, A. C., de Francisco, R., & Ionescu, B. (2020). Dimensionality reduction for EEG-based sleep stage detection: Comparison of autoencoders, principal component analysis, and factor analysis. Biomedical Engineering / Biomedizinische Technik, 65(5), 451–461. https://doi.org/10.1515/bmt-2020-0139

Thurnhofer-Hemsi, K., López-Rubio, E., Molina-Cabello, M. A., & Najarian, K. (2020). Radial basis function kernel optimization for support vector machine classifiers. arXiv preprint, arXiv:2007.08233. https://arxiv.org/abs/2007.08233

Unity Technologies. (2025). Unity real-time development platform. Retrieved February 2, 2025, from https://unity.com/

Vegas, A., Wells, B., Braum, P., Denault, A., Miller Hance, W. C., Kaufman, C., Patel, M. B., & Salvatori, M. (2025). Guidelines for performing ultrasound-guided vascular cannulation: Recommendations of the

American Society of Echocardiography. Journal of the American Society of Echocardiography, 38(2), 57–91. https://doi.org/10.1016/j.echo.2024.12.004

Velásquez-Salazar, M., Gaitán-Buitrago, M. H., & Becerra-Cardona, D. A. (2024). Habilidades técnicas en suturas evaluadas con OSATS, comparando diferentes métodos de instrucción y acompañamiento. Educación Médica, 25(3), 100904.

Verma, M., Singh, S., & Kaur, B. (2015). An overview of Bluetooth technology and its communication applications. International Journal of Current Engineering and Technology, 5(3), 1588. http://inpressco.com/category/ijcet

Wang, W., Chen, I., Hershkovich, L., Yang, J., Shetty, A., Singh, G., Jiang, Y., Kotla, A., Shang, J., Yerrabelli, R., Roghanizad, A., Shandhi, M., & Dunn, J. (2022). A systematic review of time series classification techniques used in biomedical applications. Sensors, 22, 8016. https://doi.org/10.3390/s22208016

Weiss, R., Karimijafarbigloo, S., Roggenbuck, D., & Rödiger, S. (2022). Applications of neural networks in biomedical data analysis. Biomedicines, 10(7), 1469. https://doi.org/10.3390/biomedicines10071469

Welyczko, N. (2020). Peripheral intravenous cannulation: Reducing pain and local complications. British Journal of Nursing, 29(8), S12–S19. https://doi.org/10.12968/bjon.2020.29.8.S12